

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-198086

(43) 公開日 平成9年(1997)7月31日

| (51) Int.Cl. ⁶ | 識別記号 | 庁内整理番号 | F I | 技術表示箇所 |
|---------------------------|-------|--------|--------------|---------|
| G 1 0 L 3/00 | 5 6 1 | | G 1 0 L 3/00 | 5 6 1 B |
| | 5 3 1 | | | 5 3 1 L |
| | 5 3 5 | | | 5 3 5 |

審査請求 未請求 請求項の数4 O L (全 5 頁)

(21) 出願番号 特願平8-4508

(22) 出願日 平成8年(1996)1月16日

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 松井 知子

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(72) 発明者 古井 貞照

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(74) 代理人 弁理士 草野 卓

(54) 【発明の名称】 話者認識用しきい値設定方法及びこの方法を用いた話者認識装置

(57) 【要約】

【課題】 性能を高めてゆく。

【解決手段】 モデル登録時の登録用音声(又はモデル更新時の更新用音声)が入力されると特徴パラメータの時系列に変換され(S_1)、この時系列より音声のモデル(HMM)が作成登録(又は更新)され(S_2)、詐称者を本人以外の登録話者として、登録音声(又は更新音声)を用いて話者認識実験を行い、本人棄却率と詐称者受率とが等しい値 ε_0 よりも、そのシステムの推定される誤り率の上限値 x だけ大きな詐称者受率となるしきい値 ϕ_1 と ε_0 のしきい値 ϕ_0 とから $\phi' = w\phi_1 + (1-w)\phi_0$ 、 $w = 2 / (1 + \exp(0.25t))$ ($t = 0, 1, 2, \dots$, モデル更新回数)を計算して、 ϕ をしきい値として登録(又は更新)する(S_5)。

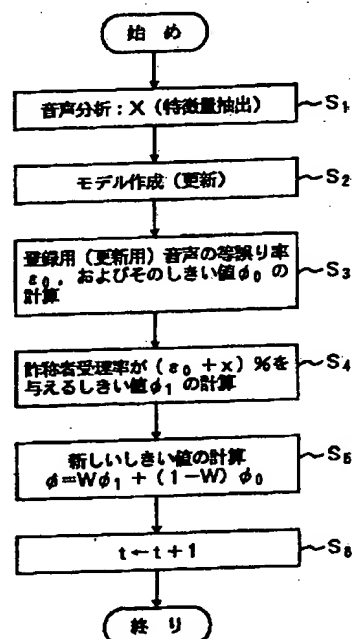


図 1

1

【 特許請求の範囲】

【請求項1】 入力音声、特徴パラメータを用いた表現形式に変換し、その表現形式による入力音声と、あらかじめ話者対応に登録された上記表現形式による音声のモデルとの類似度を求め、その類似度と話者判定用しきい値とを比較して上記入力音声を発声した話者を認識する話者認識方法における上記しきい値を設定する方法において、

各話者の上記モデル登録の際に発声された音声と上記登録モデルを用いて本人棄却率と詐称者受理率との二つの誤り率を計算し、

これら計算された二つの誤り率が等しくなるしきい値より所定値を差し引いた値に上記話者判定用しきい値を設定することを特徴とする話者認識用しきい値設定方法。

【請求項2】 上記各話者に対するモデルを更新するごとに、その更新されたモデルとその更新の際に発声された音声を用いて上記二つの誤り率を計算し、

これら計算された二つの誤り率が等しくなるしきい値より、上記所定値より小さく、かつ前回よりも小さい値を差し引いた値に上記話者判定用しきい値を更新することを特徴とする請求項1記載の話者認識用しきい値設定方法。

【請求項3】 上記所定値を、上記話者認識方法自体の誤り率の上限とほぼ等しい値とすることを特徴とする請求項1又は2記載の話者認識用しきい値設定方法。

【請求項4】 入力音声、特徴パラメータ抽出手段で特徴パラメータを用いた表現形式に変換され、この表現形式による入力音声のモデルがモデル作成手段により作成されてモデル蓄積手段に蓄積される。また、上記特徴パラメータ抽出手段で変換された表現形式の音声は類似

度計算手段で上記モデル蓄積手段内の各モデルとの類似度が計算され、これら計算された類似度は、しきい値蓄積部の本人の声とみなせる類似度の変動の範囲を示すしきい値と話者認識判定手段で比較され、類似度の方が大きければその本人の音声であり、小さければ他人の音声であると判定される話者認識装置において、

モデル更新の指示があると、上記特徴パラメータ抽出手段よりの特徴パラメータを用いた表現形式による入力音声により、これと対応する話者の上記モデル蓄積部内のモデルを更新するモデル更新手段と、

上記更新されたモデルについてその更新時の音声についての本人棄却率と詐称者受理率とを計算し、これら棄却率と受理率が等しくなるしきい値からわずかに小さな値を差し引いた値に、上記しきい値蓄積部内の対応する話者のしきい値を更新するしきい値計算手段と、を具備することを特徴とするしきい値更新を伴う話者認識装置。

【 発明の詳細な説明】

【 0 0 0 1】

【 発明の属する技術分野】 この発明は、例えば入力され

2

た音声により暗証番号の人と同一人であることを同定したりするために用いられ、入力音声を、特徴パラメータを用いた表現形式に変換し、その表現形式による入力音声と、あらかじめ話者対応に登録された上記表現形式による音声のモデルとの類似度を求めて、入力音声を発声した話者を認識する話者認識方法における、話者の判定に用いるしきい値の設定方法及びこの方法が適用された話者認識装置に関する。

【 0 0 0 2】

【 従来の技術】 図3に従来のテキスト独立形話者認識を例としたその装置の機能構成を示す。まず話者の登録をするが、各話者ごとに発声した文章などの音声（登録用音声）が入力端子11より特徴パラメータ抽出手段12に入力され、音声に含まれる特徴パラメータ（例えばケプストラム、ピッチなど）を用いた表現形式に変換され、この特徴パラメータの時系列に変換された登録用音声データから、モデル作成手段13でその音声のモデル、例えば隠れマルコフモデル（Hidden Markov Model: HMMと記す。例えば複数のガウス分布の重み付き加算で表現される）が作成される。HMMを作成する方法としては、例えば文献「松井知子、古井貞照：“VQ、離散／連続HMMによるテキスト独立形話者認識法の比較検討”、電子情報通信学会音声研究会資料、SP91-89、1991」に述べられている方法などを用いることができる。このようにして得られた話者ごとのHMMは話者対応にモデル蓄積部14に登録される。

【 0 0 0 3】 話者を認識する場合は、その話者の発声音声が入力端子11から特徴パラメータ抽出手段12に入力されて、特徴パラメータの時系列に変換され、この入力音声の特徴パラメータの時系列は類似度計算手段15でモデル蓄積部14に蓄えられた各話者のHMMとの類似度が計算され、その計算結果は、話者認識判定手段16で、しきい値蓄積部17に蓄積されている、本人の声とみなせる類似度の変動の範囲を考慮したしきい値と比較され、そのしきい値より大きければ、その入力音声は類似度計算に用いたHMMの登録話者の音声であると判定され、しきい値より小さければその他の人の音声であると判定され、この判定結果が出力される。

【 0 0 0 4】 従来においてしきい値を設定するために、本人棄却率と詐称者受理率との二つの誤り率が考慮されていた。本人棄却率は全登録用音声のうち、本人の登録用音声を用いた話者認識実験の結果から求められるもので、本人が誤って棄却される率を表し、詐称者受理率は詐称者の音声を用いた話者認識実験の結果から求められるもので、詐称者が誤って受理される率を表している。話者認識の目的によって、本人棄却率の方が詐称者受理率よりも重要であったり、またその逆であったりする。目的がはっきりしない場合には、ベイズの定理より、本人棄却率と詐称者受理率が等しい、等誤り率を与える値が最適なしきい値（等誤り率のしきい値）としていた。

50

3

図4 Aに示すように、本人棄却率を示す曲線2 1はしきい値を大きくするに従って大となる。一方、詐称者受理率を示す曲線2 2はしきい値が大きくなるに従って減少する。従来においては詐称者を本人以外の登録話者として、全登録用音声を用いて、各モデル(HMM)との類似度を計算して話者認識を行い、その際に、話者判定用のしきい値を変化させ、つまり話者認識実験を行い、図4 Aに示した本人棄却率曲線2 1と詐称者受理率曲線2 2との交点、つまり両誤り率が等しい値 ϵ_0 となるしきい値 ϕ_0 を求めてしきい値として設定し、即ち登録用音声による等誤り率のしきい値を設定していた。

【 0 0 0 5 】

【 発明が解決しようとする課題 】しかし、本人のモデルが発声内容の違い、発声変動などに対して十分に頑健でない場合は、本人のモデルとそれを作成する時に用いた本人の音声(本人の登録用音声)との類似度は、本人のモデルと本人が認識の際に発声する音声との類似度に比べて、一般に大きいものとなる。従って、本人が認識の際に発声する音声について、しきい値を変化させる本人棄却率曲線を求めると、例えば図4 Aの点線曲線2 3のように、本人の登録用音声を用いた本人棄却率曲線2 1よりも棄却率が悪くなる、つまり同一しきい値に対して棄却率が大きくなる。つまり登録用音声による等誤り率のしきい値 ϕ_0 は、認識用音声による等誤り率のしきい値 ϕ_0 よりも大きな値となり、その結果、登録用音声による等誤り率のしきい値 ϕ_0 を認識に用いた場合には本人棄却率が大きくなるという問題があった。

【 0 0 0 6 】また、本人の登録用の音声データ量はあまり多くないために、特に本人のモデルが発声内容の違い、発声変動などに対して十分に頑健でない場合が多く、本人棄却率を信頼性高く求められないことも問題であった。更に話者の声は発声の度に変動し、特に2 ~ 3 *

$$\phi = w \phi_1 + (1 - w) \phi_0$$

ここで ϕ_0 は詐称者を本人以外の登録話者として、登録用音声を用いて話者認識実験を行った時の等誤り率のしきい値、つまり最初に設定したしきい値を表し、 ϕ_1 は詐称者受理率としきい値との関係(図4 A)に基づいて、詐称者受理率が{ 等誤り率 $\epsilon_0 + x$ }%(例えば $x = 1$ %)になる値のしきい値を表す。この{ 等誤り率 $\epsilon_0 + x$ }%は、その話者認識方法の性能から推定される ※40

$$w = 2 / (1 + \exp(0.25t))$$

ここで t は話者のモデルの更新回数($t = 0, 1, 2, \dots$)を表し、この式は実験により求めた式である。この式(1)、(2)によれば、 $t = 0$ では話者認識装置を作った時、あるいは、認識すべき話者を全て新らたなものにした時であり、つまり登録用音声を用いた最初に決定するしきい値であり、等誤り率しきい値 ϕ_0 より $\Delta \phi (= \phi_0 - \phi_1)$ だけ小さいしきい値に設定される。通常、モデルの更新回数が多くなるに従って、 w が小さくなり、 $\Delta \phi$ も小さくなって ϕ_0 に近づく。なお ϕ

4

*カ月の単位で大きく変動する。この点から、高い認識性能を維持するためには定時に各話者について、音声を発声してもらい、そのモデルを更新することが望まれる。このようにモデルの更新が行われると、本人棄却率特性及び詐称者受理率特性も変化する。従ってモデル更新が行われるとしきい値も再設定することが望ましい。

【 0 0 0 7 】

【 課題を解決するための手段 】請求項1の発明の方法によれば、詐称者を本人以外の登録話者として、登録用音声を用いて話者認識実験を行った時の等誤り率を与えるしきい値から所定値だけ差し引いた値をしきい値とする、つまり等誤り率のしきい値より、高めの詐称者受理率を与える値に設定する。この高めの詐称者受理率は、等誤り率のしきい値での詐称者受理率よりも、この話者認識方法のシステム誤り率の上限程度だけ高いものである。この構成により、頑健なモデルでなくても、本人棄却率が大きくなり過ぎることはない。

【 0 0 0 8 】請求項2の発明の方法では周期的にモデルの更新を行い、その更新ごとに、その更新用音声と更新モデルとを用いて詐称者を本人以外の登録話者として、話者認識実験を行った時の等誤り率のしきい値から、前記所定値より小さくかつ更新回数に応じて小さくなる値が差し引いた値を新たなしきい値とする。この構成により、モデルの更新が行われるに従って発声内容の違いや発声変動などに対して次第に頑健になり、かつその理想的なモデルを用いた場合の認識用音声に対する理想的な等誤り率のしきい値に、前記高めの詐称者受理率を与える値から漸近してゆくことになる。

【 0 0 0 9 】前記モデル更新ごとのしきい値の更新の例を以下に示す。つまり次式に従ってしきい値 ϕ を設定する。

(1)

※詐称者受理率(そのシステム推定誤り率)の上限に対応している。 w は話者のモデルの更新に合わせて、等誤り率のしきい値にしきい値が漸近していく速度を制御するパラメータで、例えば次式のように定義することができる。

【 0 0 1 0 】

(2)

ϕ_0, ϕ_1 もモデル更新ごとに求められる。

【 0 0 1 1 】

【 発明の実施の形態 】図1にこの発明の方法の実施例における処理順を示し、図2にこの発明の装置の実施例の機能構成を図3と対応する部分に同一符号を付けて示す。図2においてこの実施例では、登録時及びモデル更新時の各入力音声の特徴パラメータの時系列が一時蓄積される特徴パラメータ時蓄積部2 5、モデル更新指示があるとモデル蓄積部1 4内のモデル更新を行うモデル更

新手段26、更に登録時及びモデル更新時にしきい値を計算してしきい値蓄積部17のしきい値を更新するしきい値計算手段27が設けられる。

【0012】入力端子11に登録用音声又は更新用音声が入力されると、図1、図2に示すように特徴パラメータ抽出手段12で特徴パラメータの時系列に変換され(S1)、登録時にはモデル作成手段13でその音声のモデルが作成され、モデル更新時には、更新用音声の特徴パラメータ時系列により、モデル蓄積部14内の対応モデルの更新が行われる(S2)。

【0013】このモデル更新のためには、各話者ごとに登録用音声、更新用音声の各特徴パラメータ時系列を保持しておき、それまでの全保持時系列と、新たに入力された更新用音声の時系列とを用いて新たにモデルを作成してモデル蓄積部14内の対応モデルを更新する。あるいは、モデルがHMMの場合、ベイズ推定により、更新用音声の特徴パラメータの時系列 X の対応話者のHMMに対する尤度 $f(X|\theta)$ と、それまでに発声された音声の特徴を反映する事前確率密度関数 $g(\theta)$ との積が最大となるHMMのパラメータベクトル θ を推定し、その θ を新たなHMMとする。

【0014】次に登録時には登録用音声、モデル更新時には更新用音声を用いて等誤り率 ϵ_0 及びそのしきい値 ϕ_0 を計算する(S3)。つまりこれら音声の特徴パラメータの時系列は特徴パラメータ時蓄積部25に一時蓄積され、これらとモデル蓄積部14内の各モデルとの類似度が類似度計算手段15でそれぞれ、計算され、これら類似度に対し、話者認識判定手段16で各種のしきい値に対し判定を行い、詐称者を本人以外の登録話者として、登録用音声(又は更新用音声)を用いた話者認識実験を行い、図4Aに示した本人棄却率曲線と詐称者受理率曲線とを求め、両誤り率が等しい誤り率 ϵ_0 と、その時のしきい値 ϕ_0 を求め。

【0015】その後、詐称者受理率が $(\epsilon_0 + x)\%$ を与えるしきい値 ϕ_1 を求め(S4)、新たなしきい値 ϕ を $w\phi_1 + (1-w)\phi_0$ を計算して求める(S5)。この新たなしきい値 ϕ を、しきい値蓄積部17内の対応話者のしきい値とする。次にモデル更新回数 t を+1して終了とする(S6)。ステップS3、S4、S5、S6はしきい値計算手段27で行われる。

【0016】図1において、 $t=0$ では登録時の登録用音声を用いたしきい値計算が行われ、その時の等誤り率のしきい値を与える誤り率 ϵ_0 よりも $x\%$ だけ高い詐称者受理率となるしきい値 ϕ_1 がしきい値として設定され、モデルが頑健でなくても、本人棄却率が大き過ぎることはない。また、モデル更新が行われるごとにその都度、その更新されたモデルについて、その更新用音声を用い、かつ詐称者を本人以外の登録話者として、話者認識実験が行われ、つまり頑健なものに近づいて来たモデルについての等誤り率しきい値に近いものとなり、かつ

wが小さくなり、その理想的に近づいた等誤り率しきい値との差が小となる、これより小さいしきい値が設定され、つまりモデル更新が繰り返される程、望ましいしきい値となる。

【0017】

【発明の効果】次にこの発明の効果を確めるための実験例を述べる。実験は、男性20名が約15カ月に渡る5つの時期(時期A、B、C、D、E)に発声した文章データ(1文章長は平均4秒)を対象とする。登録話者として男性10名、詐称者としてその他の男性10名を用いた。これらの音声を、従来から使われている特徴量、つまり、ケプストラムの短時間毎の時系列に変換する。ケプストラムは標準化周波数12kHz、フレーム長32ms、フレーム周期8ms、LPC分析(Linear Predictive Coding、線形予測分析)次数16で抽出した。登録には、時期Aに発声した10文章を用いた。更新には、1回目の更新として時期Bに発声した10文章を用い、2回目の更新として時期Cに発声した10文章を用いた。テストでは、時期D、Eに発声した5文章を1文章づつ用い、つまり時期A、B、Cによる各モデルがしきい値について各5回づつテストを行った。なお、しきい値の設定では、 $x=1\%$ とした。

【0018】この発明の効果は、テキスト独立型(例えば文献「松井知子、古井貞照:“VQ、離散/連続HMMによるテキスト独立形話者認識法の比較検討”、電子情報通信学会音声研究会資料、SP91-89、1991」)の話者認識において試した。各話者のHMMは、1状態が64個のガウス分布の重み付き加算(例えば文献「松井知子、古井貞照:“VQ、離散/連続HMMによるテキスト独立形話者認識法の比較検討”、電子情報通信学会音声研究会資料、SP91-89、1991」)で表した。

【0019】結果は、本人棄却率と詐称者受理率の平均で評価した。その結果を図4Bに示す。従来法は、詐称者を本人以外の登録話者として、全登録用音声を用いて話者認識実験を行った時の等誤り率のしきい値による結果を表す。これより、この発明方法は従来法と比べて、高い性能を示すことがわかる。これらの結果より、この発明方法は有効であることが実証された。

【図面の簡単な説明】

【図1】この発明方法の実施例における処理手順を示す流れ図。

【図2】この発明装置の実施例の機能構成を示すブロック図。

【図3】従来の話者認識装置の機能構成を示すブロック図。

【図4】Aはしきい値に対する本人棄却率及び詐称者受理率の関係を示す図、Bはこの発明の効果を説明するための実験結果を示す図である。

【 図1 】

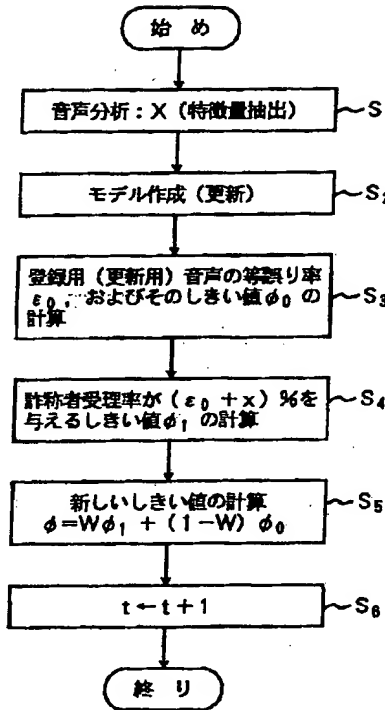


図 1

【 図3 】

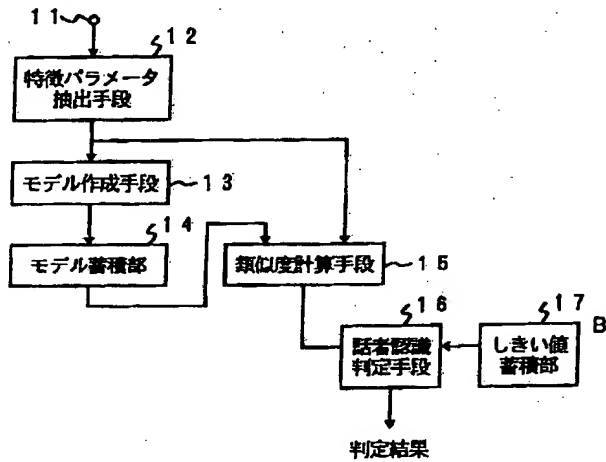


図 3

【 図2 】

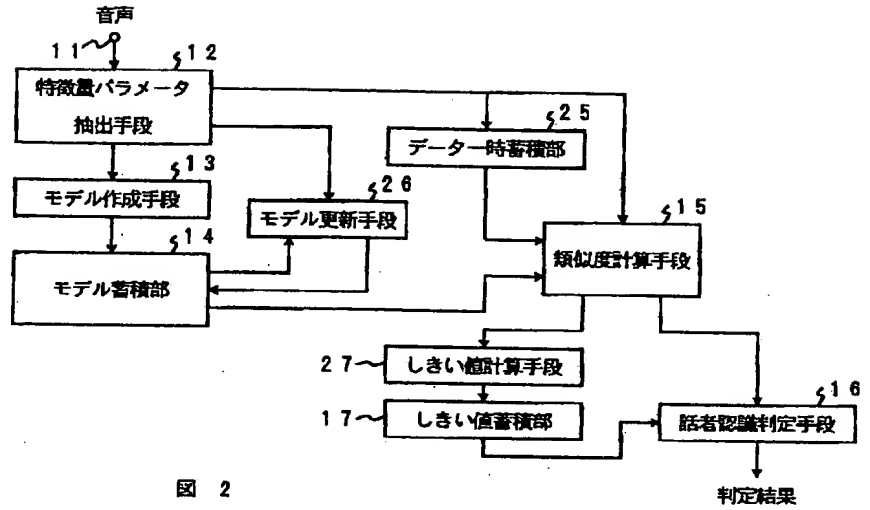
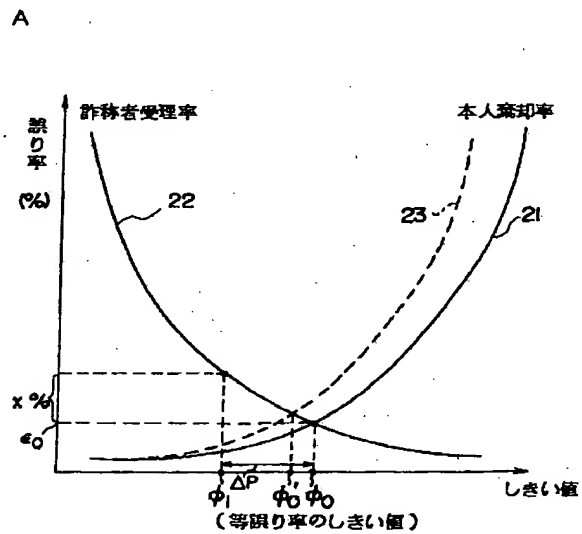


図 2

【 図4 】



| 更新の回数 | | | | | |
|-------|-----|-----|-----|-----|-----|
| 0 | | 1 | | 2 | |
| 従来法 | 本方法 | 従来法 | 本方法 | 従来法 | 本方法 |
| 11.4 | 5.1 | 5.9 | 4.4 | 7.1 | 2.2 |

図 4